REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
29-08-2013	Final Report	1-Aug-2009 - 31-Dec-2012

4. TITLE AND SUBTITLE

Final Report

Development of a Novel Technology for Label-Free DNA Sequencing

5a. CONTRACT NUMBER

W911NF-09-1-0379

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

611102

6. AUTHORS

Kurt Becker, Peiji Zhao

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

Polytechnic University - Brooklyn
Office of Sponsored Research
Polytechnic Institute of New York University
Brooklyn, NY 11201 -3840

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)

ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

56767-EL.1

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

The understanding of the sequence of the bases in a DNA molecule has many applications. Biochemical techniques are complicated, expensive, and time-consuming. Optical methods are not suitable. Here we investigate a novel sequencer concept based on a optical/quasi-optical method and electronics.

15. SUBJECT TERMS

DNA sequencing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Kurt Becker
UU	UU	UU	UU		19b. TELEPHONE NUMBER: 718-260-3608

Report Title

Final Report

ABSTRACT

The understanding of the sequence of the bases in a DNA molecule has many applications. Biochemical techniques are complicated, expensive, and time-consuming. Optical methods are not suitable. Here we investigate a novel sequencer concept based on a optical/quasi-optical method and electronics.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

TOTAL:

Number of Manuscripts:

Books

Received Paper

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Peiji Zhao	0.00	
FTE Equivalent:	**0.00**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED	Discipline
Steve Weisenholz	0.00	Electrical Engineering
FTE Equivalent:	**0.00**	
Total Number:	**1**	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	1.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:.....	1.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:.....	1.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):	1.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME PERCENT_SUPPORTED

FTE Equivalent:

Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

October 2012 report is attached.
In the remaining 3 months, the investigation of the conduction mechanism of charge in DNA molecules was completed. The IR/THz absorption studies remained unfinished.

Technology Transfer

Annual Research Progress Report for W911NF0910379

Development of a Novel Technology for Label-Free DNA Sequencing

Peiji Zhao and Kurt Becker

Department of Applied Physics

Polytechnic Institute of New York University

10/19/2012

Contents

ABSTRACT

The understanding to the sequence of the bases in a DNA molecule has widespread applications both in medical research and biology and in homeland defense with huge political and economic impacts. Biochemical techniques for DNA sequencing are complicated, expensive and time-consuming. Hence, a method that can speed up DNA sequencing with reduced time consuming and cost is very much needed in molecular biology and homeland defense. Optical methods have been widely used for the detection of the characteristics of materials because of the high sensitivity and accuracy of the methods. However, ***current optic or quasi-optic based techniques cannot provide the sequence information of the nucleotide bases in a DNA molecule. This is a long-standing unresolved and important scientific and engineering issue.*** This research project is devoted to resolve the long-standing unsolved and important scientific and engineering issue by investigating the relationship between the vibrational modes and the sequence of the bases in a DNA wire. A novel sequencer concept based on optical/quasi-optical method and electronics is proposed and studied for the propose of ultra-fast DNA sequencing and RNA expression. The results of this research will setup the basis for further engineering development of the proposed label free DNA sequencer. In the past year (Aug. 1, 2011 to Aug 30, 2012), we have finished numerical simulation of the characteristics of ssDANs for more than 6500 computing hours. This annul research report summaries the main research results in the past year based on the numerical simulation in the following aspects: Vibrational characteristics of genetic codons; Energy Structures of Genetic Codons; Influence of codon interactions on the IR/THz spectral and energy characteristics of ssDNA molecules; Crystal structures of short ssDNA molecules; Chaotic and ultrahigh frequency Rabi oscillation in coupled-double-dot system.

I. INTRODUCTION

DNA sequencing is driving genomics research and discovery. The completion of the Human Genome Project has set the stage for screening genetic mutations to identify disease genes on a genome-wide scale [1], which is a milestone for the personalized and evidence-based medicine in the 21st century. Here, accurate high-throughput DNA sequencing methods are needed to explore the complete human genome sequence for applications in clinical medicine and health care. Current electrophoresis-based sequencing technology has been widely used in the laboratories around the world although this technology is complicated, expensive, and time-consuming. To overcome the limitations of the current electrophoresis-based sequencing technology [2–5], a variety of new DNA-sequencing methods have been investigated. Such approaches include sequencing by hybridization [6], mass spectrometry-based sequencing [7– 9], sequence-specific detection of single-stranded DNA using engineered nanopores [10], and sequencing by ligation [11]. More recently, DNA sequencing by synthesis (SBS) approaches such as pyrosequencing [12], sequencing of single DNA molecules [13], and polymerase colonies [14] have also been widely explored.

As well accepted in the community, optical based techniques have been extensively used in the detection and discrimination of DNA and RNA since the 1960s because of the swiftness and accuracy of the techniques in fingerprinting the molecules. Recently, there is enhanced research interest into the absorption spectra of DNA by optic or quasi-optic techniques [15]. However, optic or quasi-optic spectral analysis as traditionally employed (i.e., performed on large ensembles or arrays of molecules) cannot provide complete sequencing information of the nucleotide bases in a DNA/RNA molecular chain. This type of spectroscopic capability has extremely important implications to bio-agent detection and identification and as such is of great interest to the general scientific community, but remains largely unresolved at this time. Here, while the measured spectral characteristics have definite physical links to the molecular dynamics and structure, it is usually not possible to assign specific spectral signatures to an exact series of base sequences even for the case of short DNA/RNA chains. For example, previous studies have shown that quasi-optical analysis can be successful in the unambiguous detection of very long-wavelength (i.e., terahertz frequency) vibrational modes that are present within biological macromolecules that are dissolved in solutions [16]. However, because many of the detectable modes are pervasive among molecules of differing base sequences even for short chains, it becomes very difficult to achieve the unique assignment of the molecular structure needed for biological detection and identification. The primary difficulty here is the lack of detailed insight into the physical relationship between genetic sequence and the resulting spectral signature. It is well accepted that the major function of DNA is to store genetic information used for protein synthetics. The genetic information is stored in the coding regions of all genetic molecules (e.g., DNA and RNA) that are known to possess sequences made up of fundamental units called codons, each of which consist of three bases [17]. This special type of structure contains genetic information that is used for guiding protein synthetics because each codon is a code for a specific amino acid. The interactions between chains of amino acids form the shape and function of proteins. Codons offer a novel approach for more detailed interpretations of spectral signatures collected from the target bio-molecules. Based on this point of view, if we know the spectral signatures of individual codons (which there are only 64) and have detailed insight into how series of codons interact to modify/augment the resulting spectra absorption characteristics, then there is hope for linking these results to the molecular chain's overall structure. Hence, whether one utilizes this information in quasi-optical sensors, or alternatively, in point sensors that infer the presence of phonon modes through their influence on electron conduction, it offers a new potential approach for identifying the bio-targets that contain this type of DNA/RNA. Therefore, insight into the dynamical characteristics of codons and the interactions among codons are of crucial importance to spectral-based sequencing of DNA. The essence of the research project is the implement of the physical characteristics of codons for label-free DNA sequencing.

II. RESEARCH PROGRESS

In the past year (Aug. 1, 2011 to Aug 30, 2012), our main research efforts focus on the understanding to the intrinsic vibrational characteristics of DNA/RNA molecules for the development of the novel sequencer. Specifically, the research focuses are in the following aspects: Vibrational characteristics of genetic codons; Energy Structures of Genetic Codons; Influence of codon interactions on the IR/THz spectral characteristics of ssDNA molecules; Chaotic and ultrahigh frequency Rabi oscillation in coupled-double-dot system.

1. Crystal Structure of Short Single Stranded DNAs

A dipole-dipole interaction model and a graphical analysis technique is presented to qualitatively explain the crystal structure of single stranded DNA (ssDNAs). Here, DNA codons, which carrying the hereditary information for protein synthetics, are the fundamental units or blocks of the crystal structures of ssDNAs. Hence, the electric interaction between the dipoles of adjacent codons determines the overall form of the ssDNA while the base paring interaction modifies the geometrical parameters. Furthermore, the H-bone interaction between bases on the backbone and the Na^+ ion mediated base pairing interaction determine the aggregation of the bases on the backbones of ssDNAs.

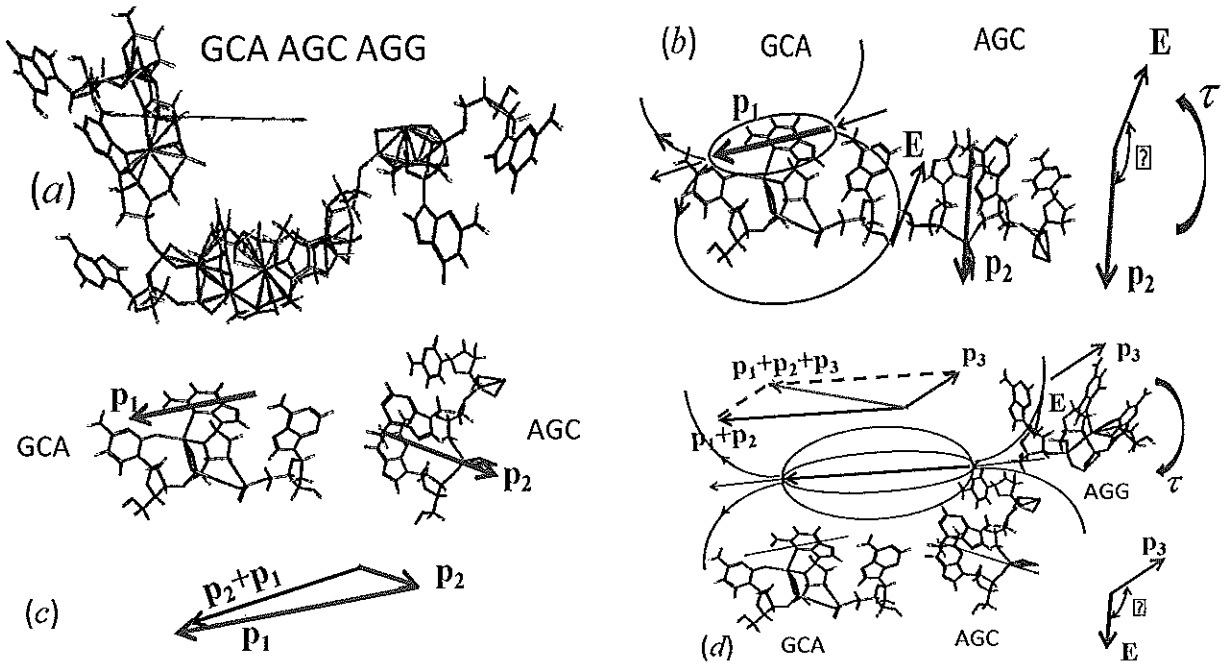


Fig.1 Model of ssDNA crystal. (a) DNA codons are the building blocks of ssDNA crystals. The base pairing interaction determines the aggregations of the bases on the backbones. (b). The interaction between of the dipoles of the codons GCA ($\mathbf{p}_1$) and AGC ($\mathbf{p}_2$) bents the backbone of the codon AGC by the torque $\boldsymbol{\tau} = \boldsymbol{p}_2 \times \boldsymbol{E}$; (c) the direction of the ssDNA GCA AGC is the sum of the individual codons; (d) the interaction between the dipole of the ssDNA GAC AGC and that of AGG ($\mathbf{p}_3$); the torque $\boldsymbol{\tau} = \boldsymbol{p}_3 \times \boldsymbol{E}$ bents the backbone of the codon AGG as shown in the figure. The direction of the torque is the sum of

the individual codons which approximates that from the ab initio calculation.

2. Vibrational Characteristics of Genetic Codons

The vibrational characteristics of DNA molecules is critical to the study of the relationship between the IR/THz wave modulated current through the molecule and the sequence of the nucleotides in the molecule. To this connection, we have theoretically investigated the vibrational characteristics of all genetic codons for the significance of the codons in the storage of bio-information in DNA molecules. This study reveals a number of important conclusions that have scientific relevance to spectroscopic characterization of DNA molecules and the sequences of the nucleotide bases in the molecules. The main results are listed below.

1) The absorption spectra of DNA codons in infrared regime shows two group feature, that is, there is a spectral forbidden region in the spectra from 2000 cm^{-1} to 3000 cm^{-1} as shown in the figure above. For the vibrational modes in the high frequency end of the spectra, the vibrational modes are localized. In the lower frequency end (100 cm^{-1} to 2000 cm^{-1}), the vibrational modes changes from quasi-localized modes to collective vibrational modes. Here, there is no one-to-one relationship between the vibrational modes and the spectral lines in this frequency region. It is believe that the vibrational modes in the lower frequency end in the region (THz frequency regime) are related to some fundamental biological processes. However, the vibrational modes in this region are not suitable to the propose of optic-based label-free DNA sequencing since there is no one-to-one relationship between the positions of the nucleotides in the molecules and the spectral lines of the molecules.

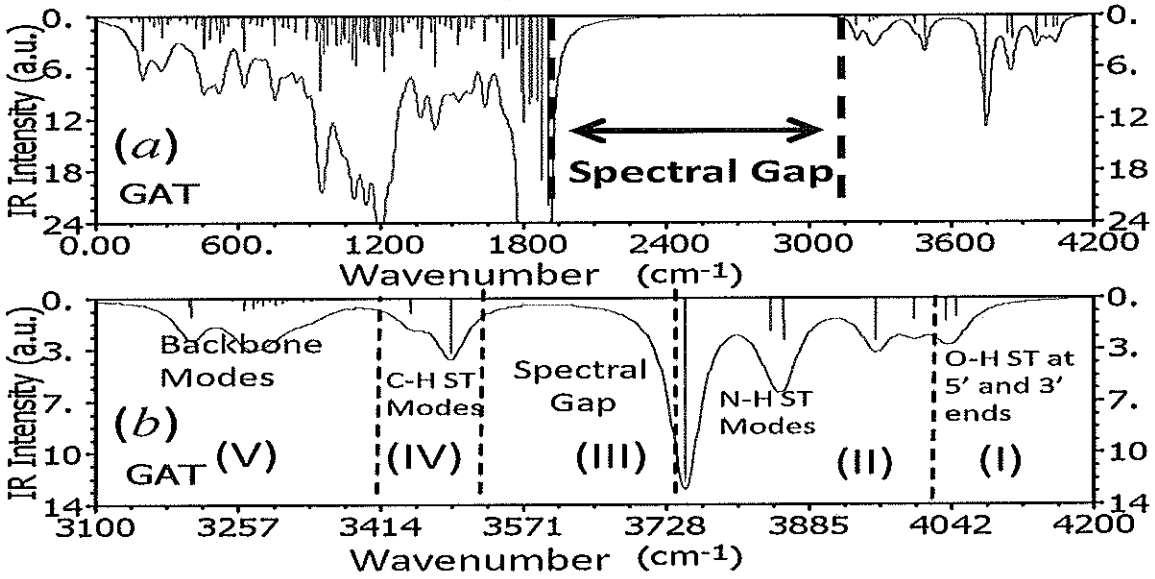


Figure 1 Typical spectral characteristics of DNA codons.

2) We have found that the spectra of the all DNA codons in gas phase in the high frequency end of middle infrared regime can be classified into five semi-distinct spectral sub-regimes,

namely: a) two end regions, two sequence identifier regions, and one region that separates the sequence identifier region; b) The spectra in the high-frequency end-regions arise primarily from vibrational modes associated with the O-H bonds at the 5'end or 3' end of the codons and the spectra in the low-frequency end-regions are formed primarily by the vibrational modes of the backbone; c) In the lower-frequency side of the regime, the sequence-identifier region is composed of the C-H bond stretch vibrations in the planes of the corresponding DNA bases. In the higher-frequency side, sequence-identifier region is composed of the N-H bond stretch vibrations in the planes of the corresponding DNA bases. In the higher frequency side, the sequence-identifier region almost exclusively contains vibrational modes due to coupling between the G nucleotides and other bases. All those of the conclusions have been theoretically verified by first principle based numerical simulations in HF approximations. The most important conclusion is: All vibrational modes in the first sequence identifier region (N-H bond vibrational modes) and in the second sequence identifier region (C-H bond vibrational modes) are localized on the corresponding nucleotides. ***This property of the modes setup a one-to-one relationship, which is crucially important to optic-based label-free DNA sequencing techniques. This conclusion also setups the foundation for the resolution of the above mentioned long-standing resolved issue.***

3) **Interactions between the Bases on the Spectral Characteristics of the IR Absorptions : Directional characteristics of the interaction between nucleotide bases** In addition to the characteristics of the vibrational modes mentioned above, all the vibrational modes in sequence identifier regions (high frequency end in the region) shows the following features: a) The spectral lines at high frequency end of the spectrum show "two group feature"; b) There is a frequency forbidden region in the region with some exceptions; c) The number of spectral lines increases while the nucleotide at the third place of a codon changes from T to G.

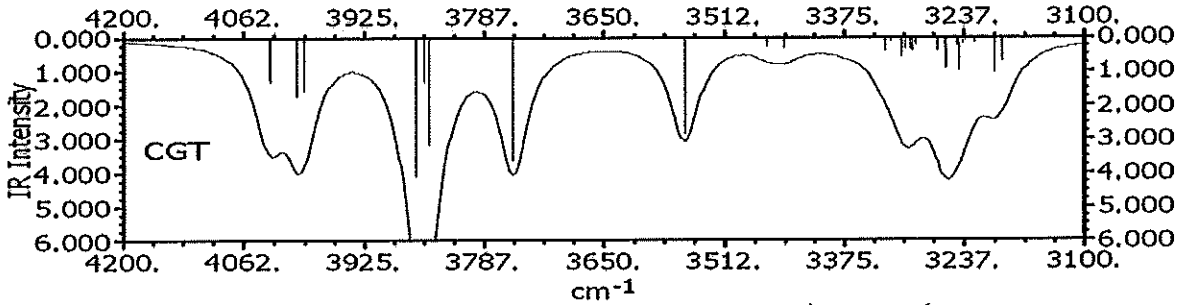


Figure 2 Absorption spectrum of DNA codon CGT in frequency domain 3100 cm^{-1} ~ 4200 cm^{-1}.

The spectral characteristics of the codons show directional characteristics because of the interaction between nucleotide bases as shown in the figures below.

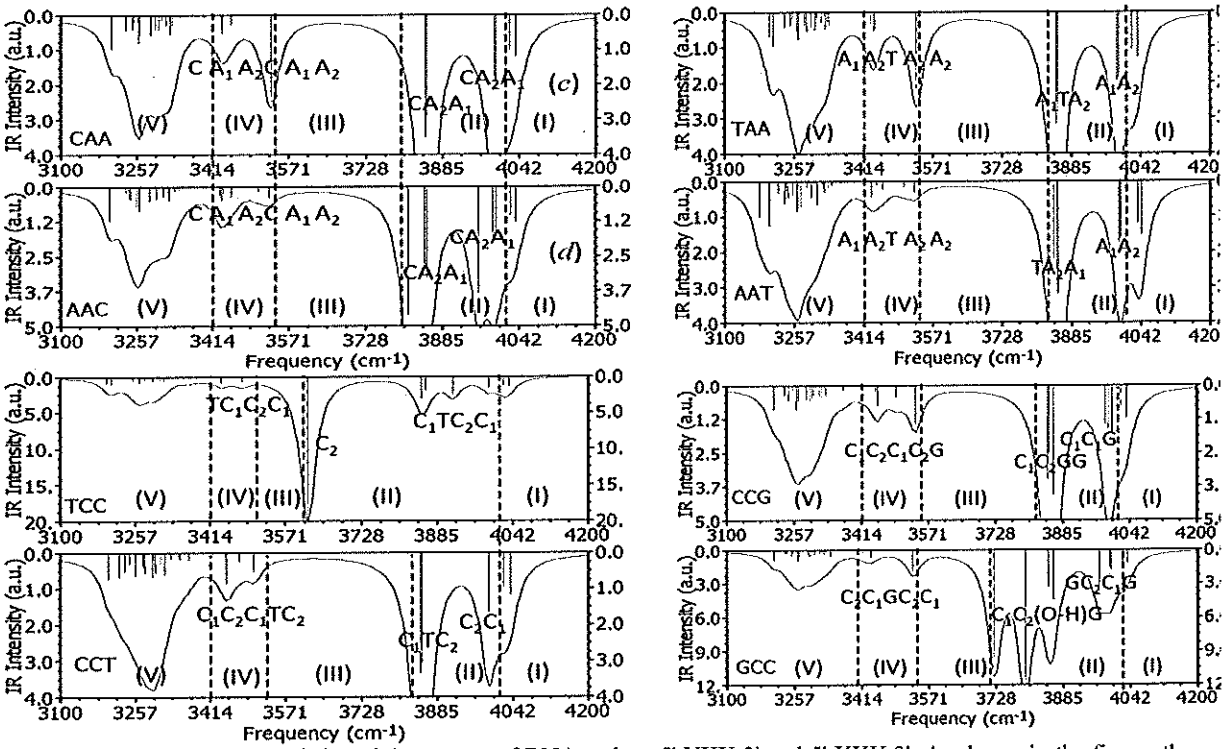
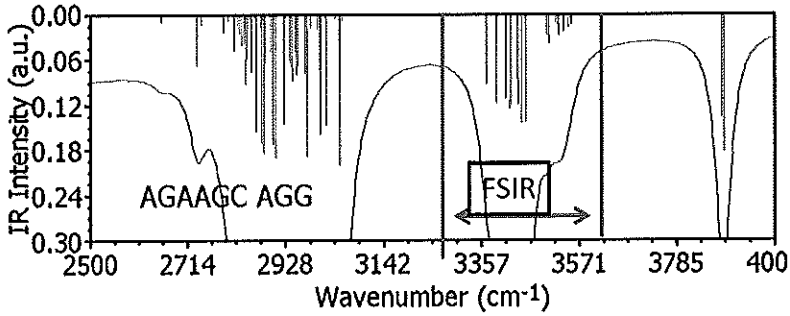


Figure 3 Directional characteristics of the spectra of DNA codons 5'-YXX-3' and 5'-XXY-3'. As shown in the figure, the spectra of the codons with base Y at the 5' ends is greatly different from those of codons with Y bases at the 3' end. The difference is caused by the differences of the conformations of the codons.

4) Relation between the number of spectral lines and the nucleotide bases

One of the most important conclusions of the research is the relationship between the numbers of spectral lines in the first sequence identifier regions of the spectra of DNA codons as shown in the figure below. Here, as shown in the figure, one T base contributes

3553.9	3548.3	3541.2	3537.5	3529.5	3527.4	3521.9	3504.9	3499.5	3454.8	3445.6
A2	A1	A4	G3	G1	G4	A2	G2	C	A3	A1
3438.7	3436.7	3423.8	3420.3	3411.9	3391.7	3384.3	3384.0	3382.2	3373.7	3370.2
A4	A2	G3	G4	G1	C	G1	G2	G3	G4	G2



$$N = N_T + 2(N_A + N_C) + 3N_C$$

$$N_T = 0, N_A = 3, N_C = 2,$$

$$N_G = 4$$

$$N = 22$$

one spectral lines, one A base or C base contributes two spectral lines, and one G base contributes three spectral lines. Hence, if the number of T bases in a ssDNA is N_T, the

number of C bases or A bases in a ssDNA is N_C or N_A, and the number of G bases in a ssDNA is N_G, the total number of spectral lines in first frequency identifier region is

$$N = N_T + 2(N_C+N_A) + N_G$$

3. Sequence dependence of the energy structures of DNA codons

It is widely believed that next generation electronic chips made out of molecules encoding all the information for the circuits. Electronic devices that naturally interface with biological tissues for development of novel medical diagnostic tools and for functional electronic devices [18]. It should be noted that all the technologies motioned above, DNA based information processing technologies and DNA sequence specific detection and discrimination technologies, are all related to the contact between DNA molecules and electrodes for information input and or readout. Hence, the design of the electrodes of the devices is one of the crucially important issues for the development of ultra-sensitive electrodes for information input/output and for detection of bio-agents. Hence, the interaction between DNA molecules and the electrodes is one of the key factors in the design of the electrodes.

It is well accepted that the genetic information is stored in the coding regions of all genetic molecules (e.g., DNA and RNA) that are known to possess sequences made up of fundamental units called codons [17]. Here, as we have discussed before, codons offer a novel approach for more detailed interpretations of spectral signatures collected from the target bio-molecules. Based on this point of view, if we know the energy structure of individual codons and have detailed insight into how series of codons interact to modify/augment the resulting energy structures, then there is hope for linking these results to the molecular chain's overall structure. Hence, it offers a new potential approach for identifying the energy structure of DNAs that contain this type of DNA/RNA. Naturally, the energy structures of the probe DNA could greatly influence the transfer of charges from DNA molecules to the electrodes, thereby affecting the sensitivity of the biosensing devices. Obviously, the energy structures of DNA codons could influence those of the probe DNA's. Therefore, insight into the energy level structures of codons and the interactions among codons are of crucial importance to design of electrodes of nucleic acid based molecular devices.

Previous calculation results show that the sequence of nucleotide bases in a DNA codon can greatly influence the HOMO and LUMO energies of the codons. If the third nucleotide in a codon is changed from Pyrimidine bases to purine bases, the change of the energy difference between the LUMO and the HOMO energies of the codons is big ranging from 0.01 eV to 0.08 eV which shows a strong sequence dependence. If the codons have the similar sequence, like GGA and AGG, the interaction of the codons produce a large number of energy levels next to the zero energy of the system. The change in sequence of the second codon, for example GGA and AGT, could greatly influence the energy structure of the system. Specifically, the major conclusions of the research are listed below.

1) The HOMO and LOMO states of codons.

The calculation results show that the HOMO and LUMO states of the codons are generally have the states shown in the figure below. Clearly,

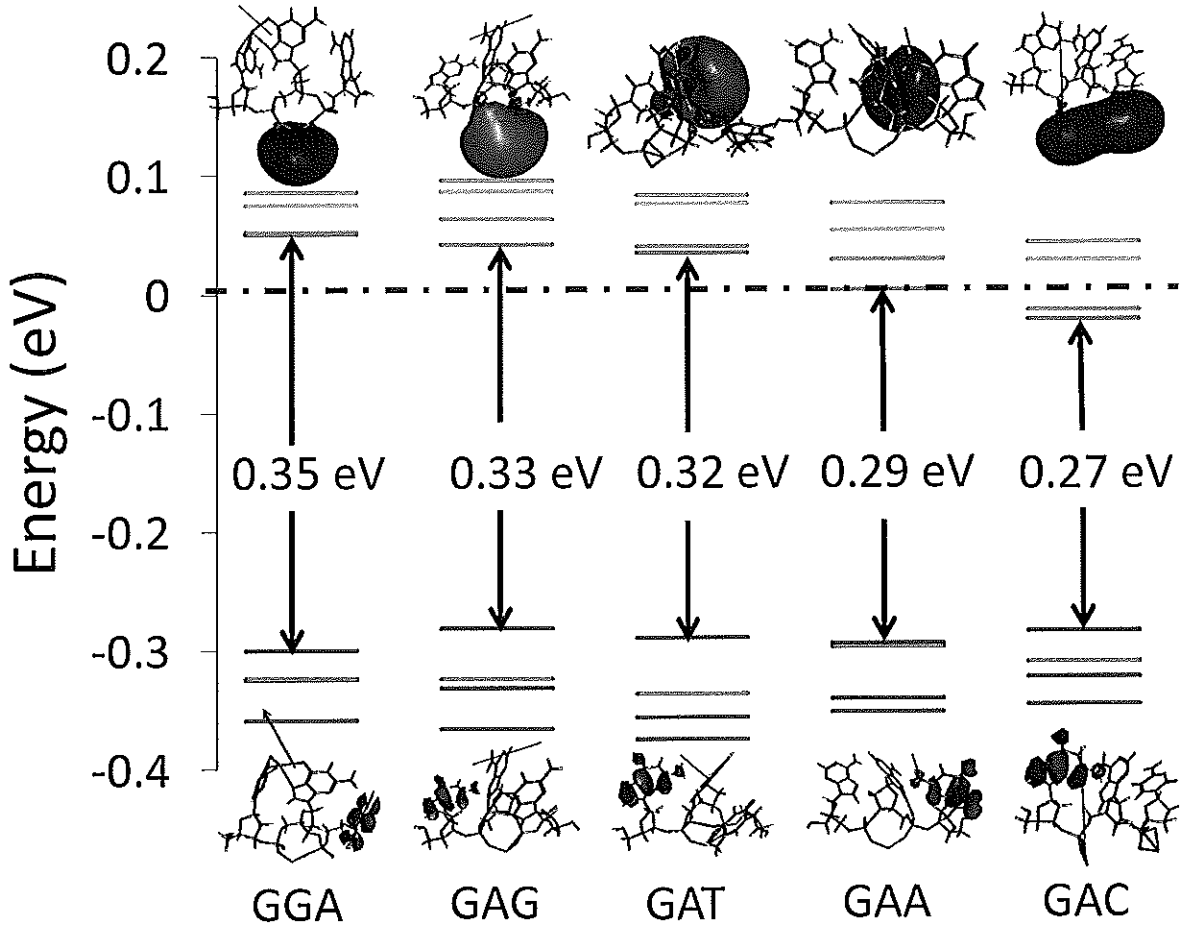


Figure 4 The HOME and LUMO states of DNA codons.

the HOME states of the codons are always located on the nucleotide bases. Specifically, the HOME states are located on A base if the third letter of the codons is A. In contrast, the states are located on the G base of the third letter is not A base. This figure shows clearly the sequence dependence of the HOMO states of the codons. Generally, the LUMO states are located in the backbones of the codons. The values of the energies of the states reduce if there are portions of the states on the nucleotide bases. Moreover, the LUMO energy becomes negative if the state is expanded over the whole backbone which greatly reduce the energy difference between the LUMO state and the HOMO state. It should be noted that the positions of the LUMO and HOMO states will influence the state of the interacting codons.

2) Influence of the sequence of nucleotide bases on the energies of the codons

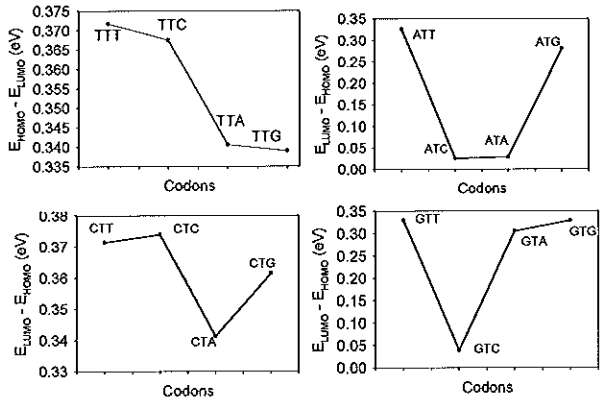


Figure 5 Variation of E_{LUMO}-E_{HOMO} with the variation of the sequence of nucleotides in DNA codons

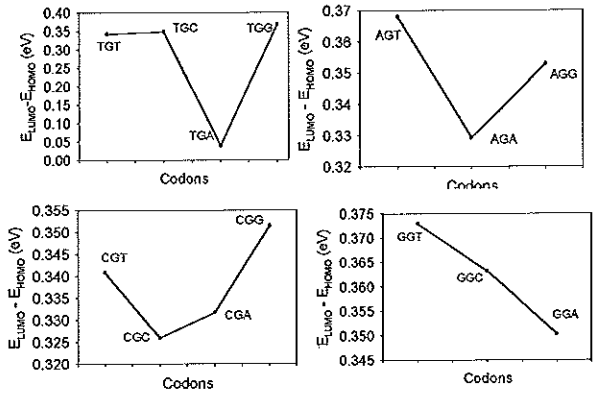


Figure 6 Variation of E_{LUMO}-E_{HOMO} with the variation of the sequence of nucleotides in DNA codons

For codon 5'-TTX-3', The difference of the energies of LUMO and HOMO reduces dramatically (~ 0.028 eV) when the third nucleotide C -> A; The change of the energy difference when X changes from Purine Bases to Pyrimidine Bases (C,T) is small. For codon 5'-CTX-3', the above stated feature remains unchanged. However, the change of the energy difference in purine bases is large; For codon 5'-ATX-3', the change of the energy difference in purine bases or in pyrimidine bases is big. There is no change of the difference of the energy when X changes from pryimidine base C to purine base A; For codon 5'-GTX-3', the energy difference inside purine bases is small. It is big inside pryimidine bases and when the base changes from C to A.

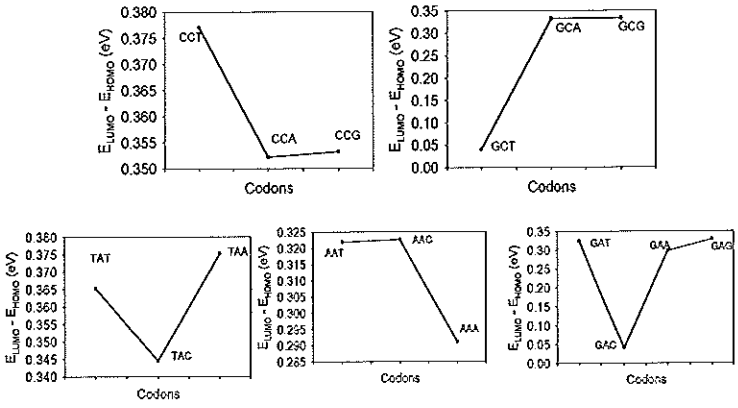


Figure 7 Variation of E_{LUMO}-E_{HOMO} with the variation of the sequence of nucleotides in DNA codons.

For codon 5'-TGX-3', In contrast to 5'-GAX-3', the change of the energy difference is small when X is T or C. The change is big when X changes from Pyrimidine Bases to purine bases or when X changes within pyrimidine bases. For codon 5'-CGX-3', 5'-AGX-3', 5'-GGX-3', the GA and GG pairs could increase the energy difference. In contrast, the GT and GC pairs would like to reduce the difference

As shown in the figure, the change of the third base inside pruine bases leads to small energy change. The energy difference is great when the third bases change from pruine bases to pyrimine bases, for example, from T to A. For codon 5'-TAX-3', the energy difference is great when the third bases change from pruine bases to pyrimine bases, for example, from C to A. The difference is also great when the third base changes from T to C; For 5'-AAX-3', the Energy difference is small when the third nucleotides are pyrimine (T, C). The change is great when the

base changes from pyrimine to pruine base; For 5'-GAX-3',the change is small when X are pruine (A and G) and the change is big when X are pyrimine bases (T, C) and when X changes from prumine (A) bases to pyrimine base (C)

4. Influence of Codon Interactions on the Energy Structures of ssDNAs

1) Origin of the Energy States of ssDNAs

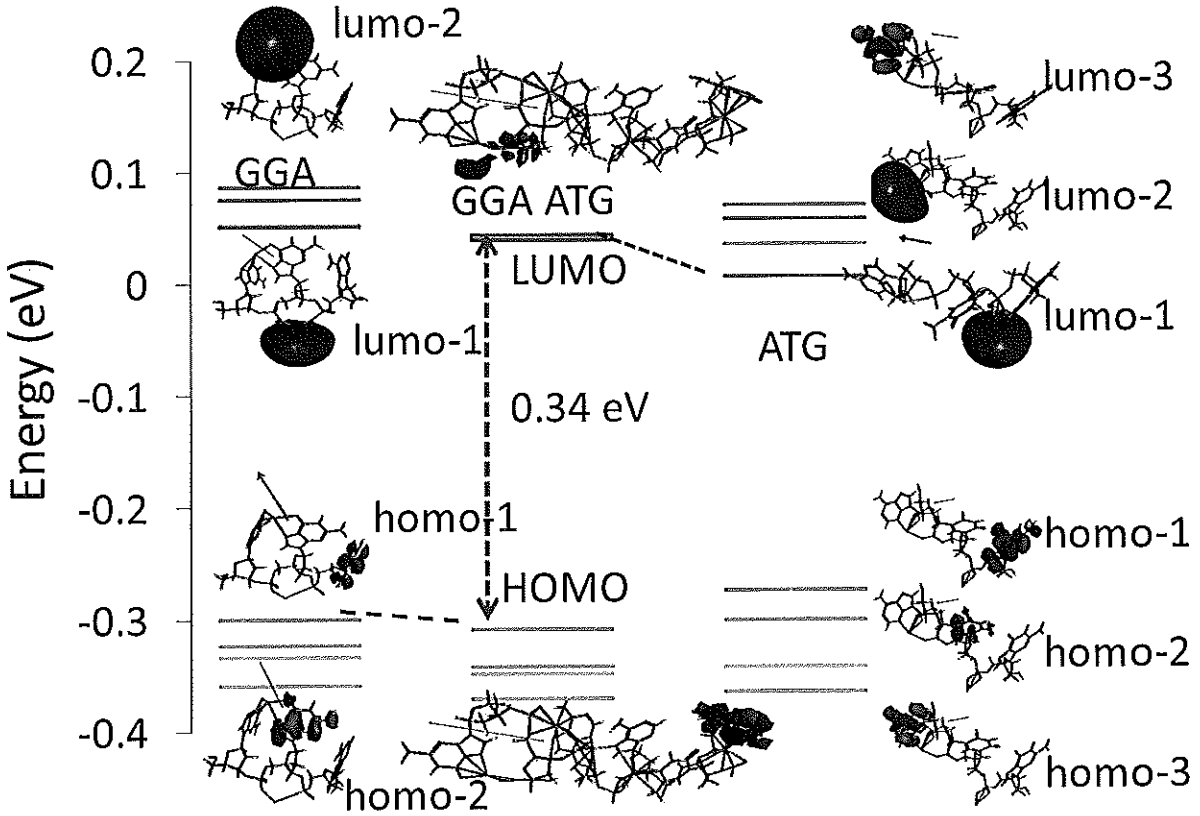


Figure 8 Origin of the energy structure of ssDNA 5'-GGA ATG-3'. As clearly shown in the figure, the LUMO state of the ssDNA comes from the interaction between the LUMO-1(2) states of the codon 5'-GGA-3' and the LUMO-1(2,3) of the codon 5'-ATG-3'. The HOME state of the ssDNA 5'-GGA ATG-3' is located at the G base next to the 3' end of the ssDNA, meaning that the energy level comes from the interaction among the HOMO-1 state of 5'-GGA-3' and the HOMO-1 (2, 3) states of the codon 5'-ATG-3'.

As the building block of genetic molecules, the interactions among the DNA codons will determine the energy states of the ssDNAs which are composed of the codons. Hence, the understanding to the origin of energy states of ssDNA is an important biophysical issue. It is also of importance to the detection of ssDNAs by modulation of the charge transfer inside of the molecules and to the design of ssDNA based optical devices. Our numerical simulation results show that the energy structures of ssDNAs can be explained in terms of multilevel interaction model. The figure above illustrates an example and explains the model. Here, as we can see that the HOMO state of the ssDNA 5'-GGA ATG-3' is localized at G base next to the 3'

end of the molecule. Hence, the HOMO-1 state of the codon 5'-ATG-3' should contribute the most to the state. However, it is important to involve the states of HOMO-1 of 5'-GGA-3' and those of HOMO-2 and HOMO-3 of 5'-ATG-3' so that the HOMO-1 state of the codon 5'-GGA-3' can transfer to the G base next to the 3' end of the ssDNA 5'-GGA ATG-3'. Hence, the HOME state of the ssDNA 5'-GGA ATG-3' mainly comes from the interaction among the HOMO-1 of 5'-GGA-3' and those of HOMO-1, HOMO-2, and HOMO-3 of 5'-ATG-3'. Similar conclusion for the LUMO state of the ssDNA 5'-GGA ATG-3' can be obtained by analyzing the states of the codons in terms of the multilevel interaction model: the LUMO state mainly comes from the interactions among the LUMO-1 and LUMO-2 states both from codon 5'-GGA-3' and 5'-ATG-3'.

2) Sequence dependence of the energy structure of ssDNAs

Since **DNA** molecules are composed of nucleotide bases, the sequences of the nucleotide bases in the molecules are one of the important factors for the determination of the energy states of ssDNAs. Here, the interaction between the codons is the key factor for the understanding to the energy levels of ssDNA molecules. To talk about the interaction between the energy levels from different codons, we should define when the interaction is strong and when the interaction is small. Physically, the interaction between the states is described by the overlap of the wavefunctions of the states from different codons. Hence, if the states are localized in spatial positions next to each other, the interaction is stronger. Otherwise, the interaction is smaller. Based on this definition, we can analyze the sequence dependence of the energy states of ssDNAs as well as the relation between the relation and the interaction between codons.

To discuss the sequence dependence, the energy structure of the ssDNA 5'-GGA AGG-3' is used as the reference in the following discussion of the relation. Here, the sequence of the first codon 5'-GGA-3' is not changed. While the third base of the second codon is changed from G to A, as shown in the figure, the LUMO state of the ssDNA 5'-GGA AGA-3' is located in the middle of the ssDNA. Considering the separations between the lumo-1(2) of 5'-GGA-3' and that of lumo-1 (2, 3) of the codon AGG is almost the same and the states are localized in the middles of the codons, the LUMO state of the system is mainly from the multilevel interaction which places the state in the center of the ssDNA 5'-GGA AGA-3'. This interaction does not reduce the energy of the LUMOS state too much. The analysis of the HOMO state is similar to that discussed above. As shown in the figure, the LUMO state of the ssDNA 5'-GGA AGA-3' is localized next to the 3' end of the molecule. Obviously, the state is mainly from the states of HOMO-2(3) of 5'-AGA-3' although there is a interaction between the HOMO-1 states of the codons. Hence, the energy of the HOMO state of the ssDNA 5'-GGA AGA-3' is not away too much from those of the codons. All those factors lead to that the ssDNA 5'-GGA AGA-3' has a larger energy difference between the LUMO state and the HOMO state of the molecule than that of the ssDNA 5'-GGA AGG-3'.

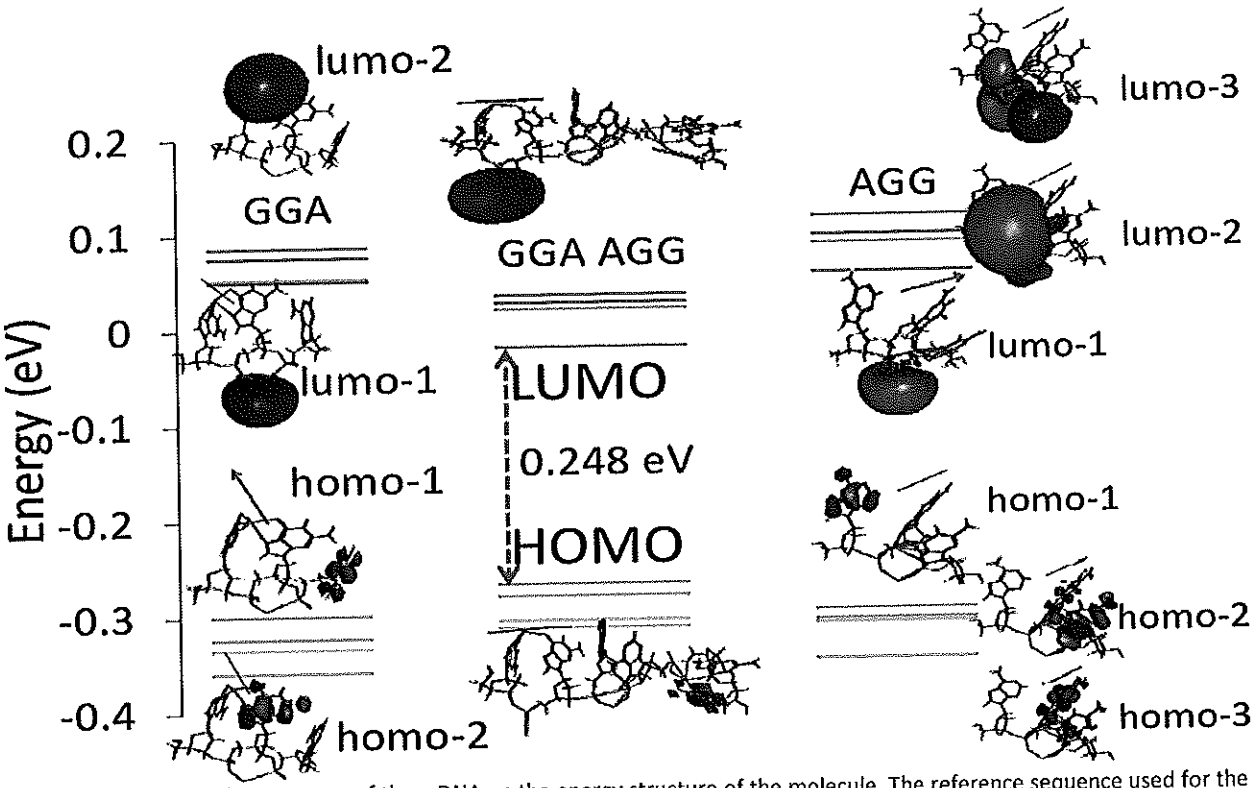


Figure 9 Influence of the sequence of the ssDNA on the energy structure of the molecule. The reference sequence used for the discussion is GGA AGG.

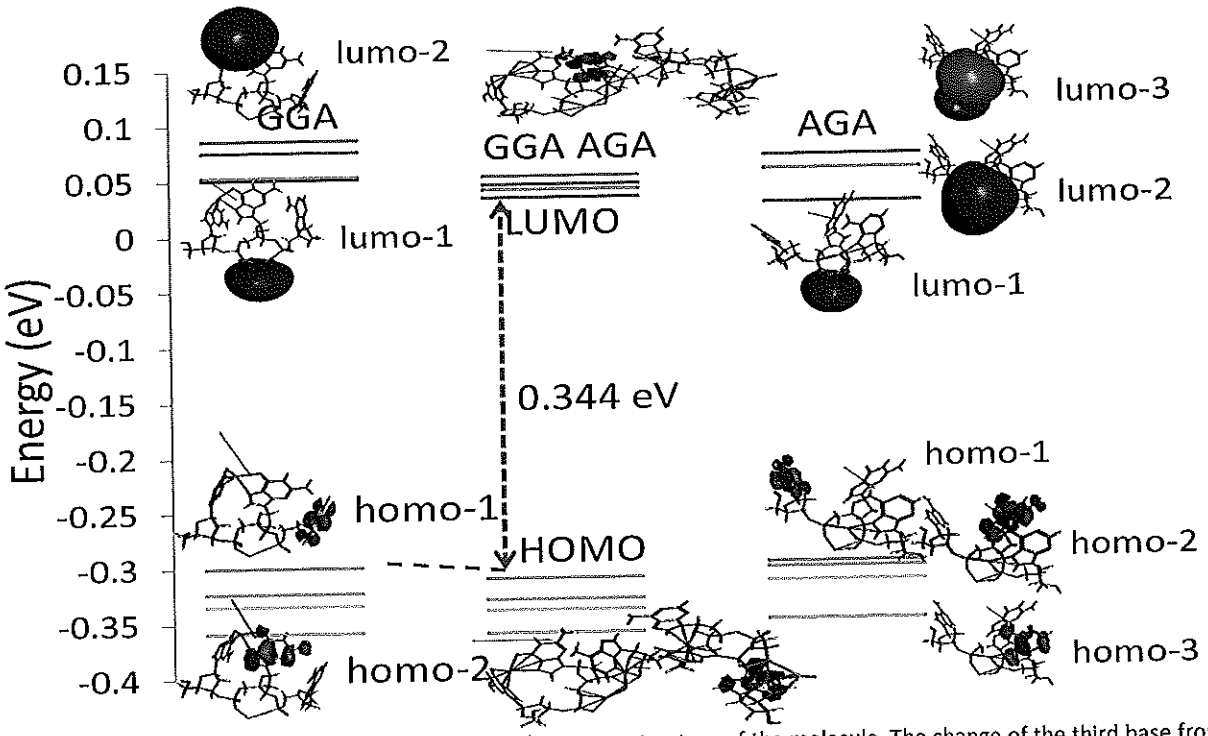


Figure 10 Influence of the sequence of the ssDNA on the energy structure of the molecule. The change of the third base from G to A in the codon AGG could lead to the increase of the energy gap between the LUMO and the HOMO states of the ssDNA 5'-GGA AGA-3'.

For the ssDNA 5'-GGA AGT-3', where the third base is changed from G to T as shown in the figure, the energy difference between of the LUMO state and the HOMO state of the molecule is smaller than that of the ssDNA 5'-GGA AGA-3' but is greater than that of 5'-GGA AGG-3'. As shown in the figure, the LUMO states of the codon 5'-AGT-3' is almost the same as those of 5'-AGA-3'. However, the HOMO states are little bit different from those of 5'-AGA-3'. Here, the HOMO state of 5'-GGA AGT-3' is in the A base of the codon 5'-AGT-3', which means that the interaction between the codons here is stronger than that between the codons 5'-GGA-3' and 5'-AGA-3'. Obviously, the interaction will push the HOME of 5'-GGA AGT-3' to a higher place which leads to s smaller energy gap between the LUMO and the HOME states of the molecule than that of the ssDNA 5'-GGA AGA-3'.

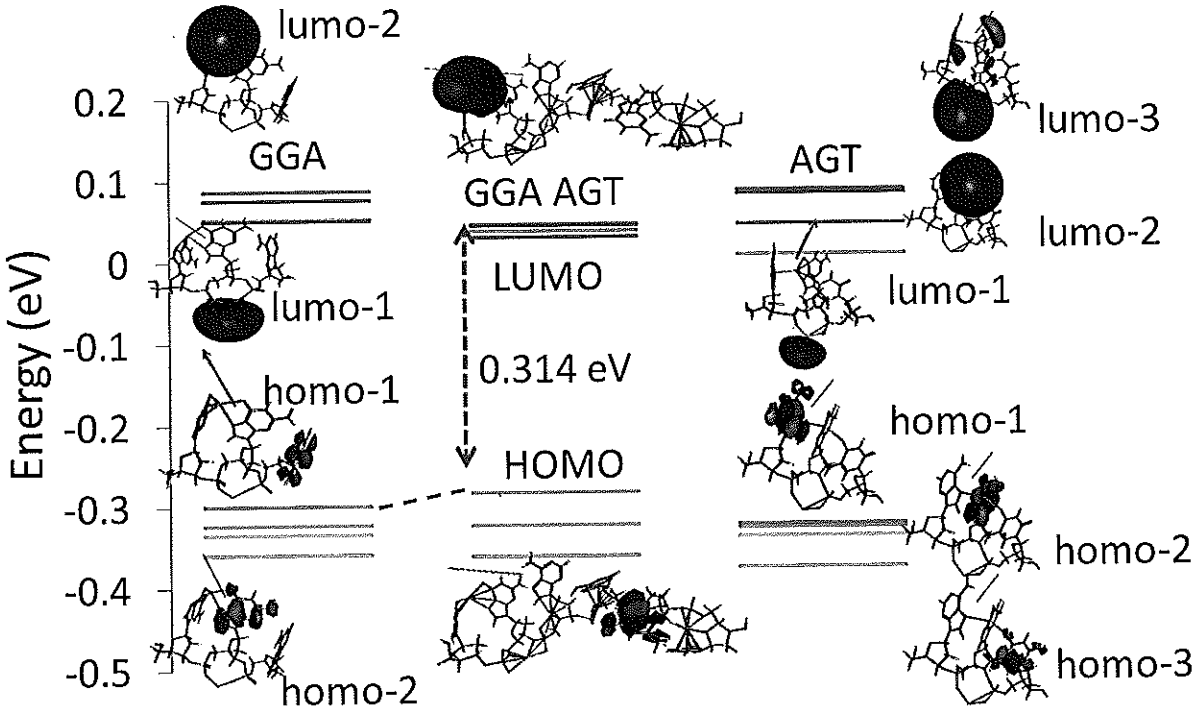


Figure 11 Influence of the sequence of the ssDNA on the energy structure of the molecule. The change of the third base from G to T in the codon AGG could lead to the increase of the energy gap between the LUMO and the HOMO states of the ssDNA 5'-GGA AGT-3'. However, the increase is smaller than that when the G base is replaced by an A base.

Considering the energy levels of the ssDNA when the second base changing from G to T. Here, as shown in the figure, the interaction is mainly determined by the LUMO-1(2) of the codon 5'-GGA-3' and the LUMO-2(3) states of the codon 5'-ATG-3'. Hence, the interaction leads to that the almost continuous LUMO states of the ssDNA 5'-GGA ATG-3'. Furthermore, the weaker interaction between the LUMO-1 states of the codons places the energy of the LUMO of the ssDNA 5'-GGA ATG-3' on a place closer to the LUMO-1 of the codon 5'-GGA-3' because the HOMO state of the ssDNA 5'-GGA ATG-3' is in the codon 5'-GGA-3' . From the spatial distribution of the HOMO states of the codons we can see that the HOMO state of the ssDNA 5'-GGA ATG-3' is mainly determined by the lower states of the HOMO states of the codons which

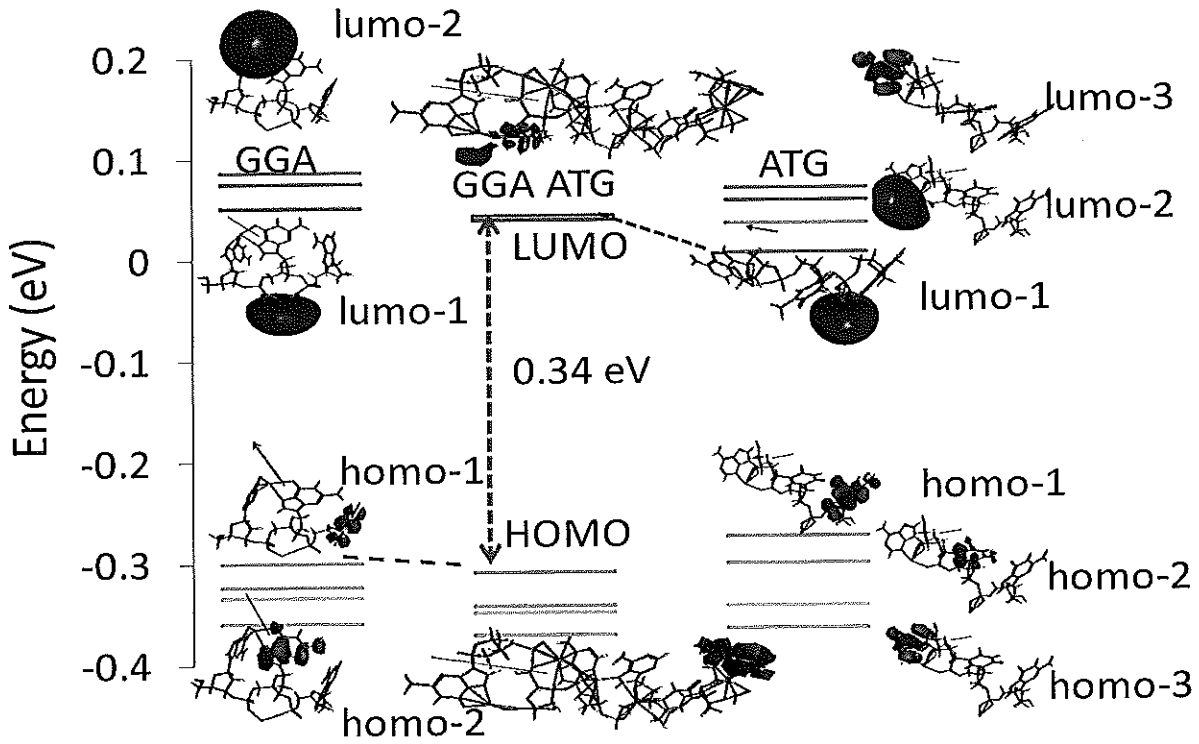


Figure 12 Influence of the sequence of the ssDNA on the energy structure of the molecule. The change of the second base from G to T in the codon AGG could lead to the increase of the energy gap between the LUMO and the HOMO states of the ssDNA 5'-GGA ATG-3', which is similar to that in the ssDNA 5'-GGA AGA-3'.

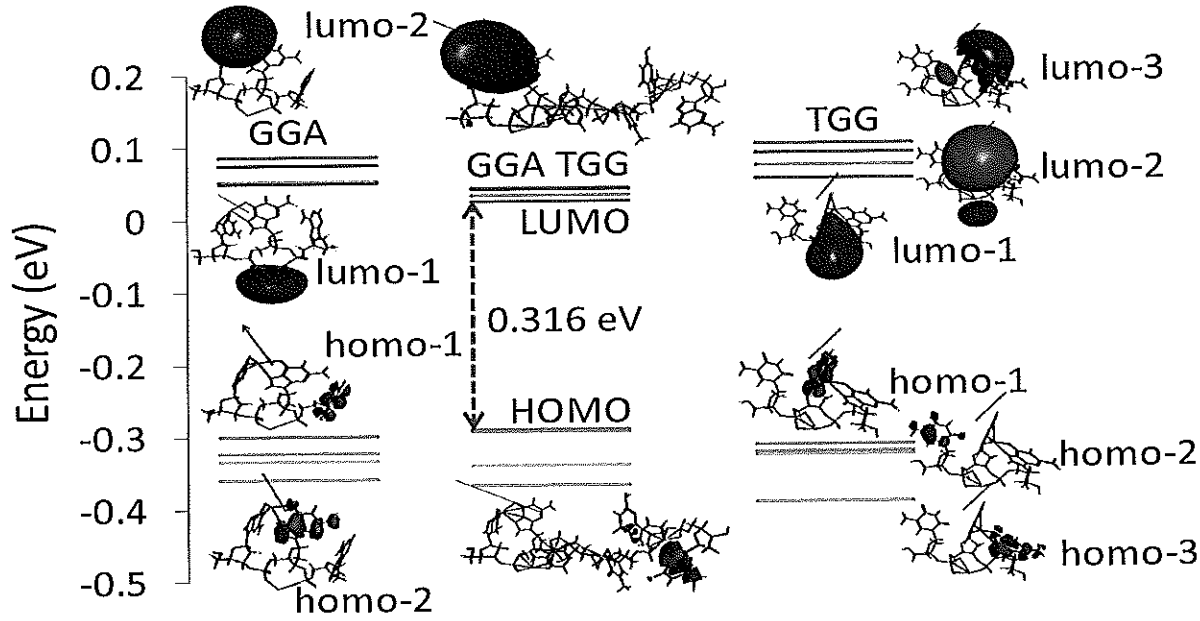


Figure 13 Influence of the sequence of the ssDNA on the energy structure of the molecule. The change of the first base from A to T in the codon AGG could lead to the increase of the energy gap between the LUMO and the HOMO states of the ssDNA 5'-GGA TTG-3'. The change is smaller than that for 5'-GGA ATG-3' or 5'-GGA AGT-3'

lowers the energy of the HOMO of the ssDNA 5'-GGA ATG-3' and leads to a larger energy difference between the LUMO and the HOMO energy states of the ssDNA 5'-GGA ATG-3'.

When the first base of the codon 5'-AGG-3' is changed from A base to T base, the energy difference between the LUMO and the HOMO states of the ssDNA is increased. Here, the LUMO-1 states of the codons have the similar energy. Hence, the stronger interaction between the energy level places the LUMO energy at the position lower than that of 5'-GGA-3'. Obviously, the interaction is weaker because the LUMO-1 states of the codons are localized in the center of the codons. For the HOMO states of the codons, the interaction between the HOMO-1 state of 5'-GGA-3' and that of the 5'-TGG-3' is greater as shown by the HOMO state of 5'-GGA TGG-3' which rises the energy of the HOMO state. These results reduce the value of the energy difference between the LUMO state and the HOMO state of the ssDNA.

5. Influence of water molecules on the absorption characteristics of DNA codons

Hydration of DNA plays important role in its structure, conformation, and function. X-ray crystallography, NMR, dielectric relaxation, and molecular dynamics simulation studies have shown that a significant amount of water molecules are bound to DNA[19]. For example, measurements of dielectric relaxation caused by water molecules bound to DNA in mixed water-ethanol solutions have found that [20] water molecules per nucleotide are present in B-DNA, but only [21] water molecules are bound in A-DNA. The study also suggested that a structural transition of poly(dG-dC)poly(dGdC) DNA from its B to Z form takes place on the removal of the bound water molecules, preferentially from the phosphate groups. Obviously, because the water molecules may change the conformations of the ssDNAs, it is necessary to understand the way that how do the water molecules change the relationship between the spectra and the nucleotide bases in the molecules. In this study, the codon 5'-GAT-3' is employed as an example for the investigation.

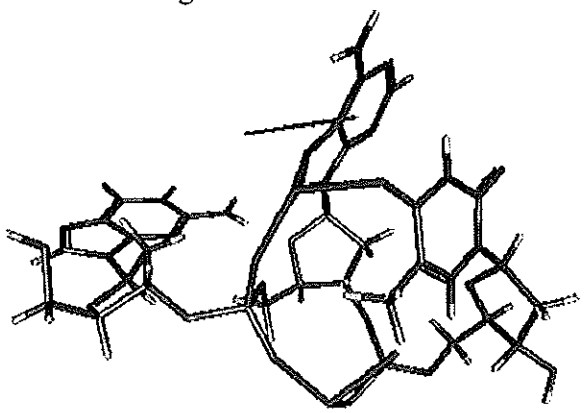


Figure 14 Crystal structure of DNA codon GAT.

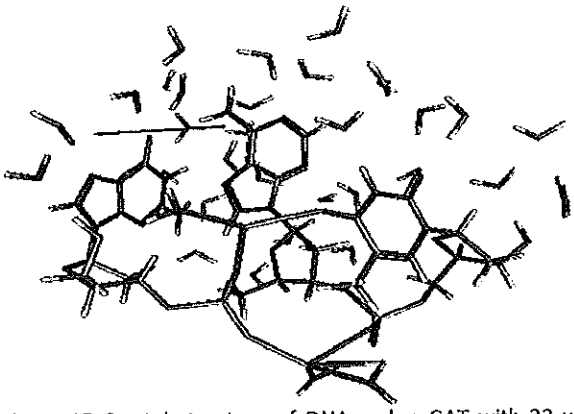


Figure 15 Crystal structure of DNA codon GAT with 33 water molecules.

The above figures shows that the crystal structures of the codon GAT with/without water molecules are greatly different. Specifically, the directions of the dipole moments for both cases are different, almost in opposite directions. Obviously, this will lead to different vibrational modes and different IR absorption modes. Moreover, as shown in the figure below, the energy structure of the codon shows strong environment related features.

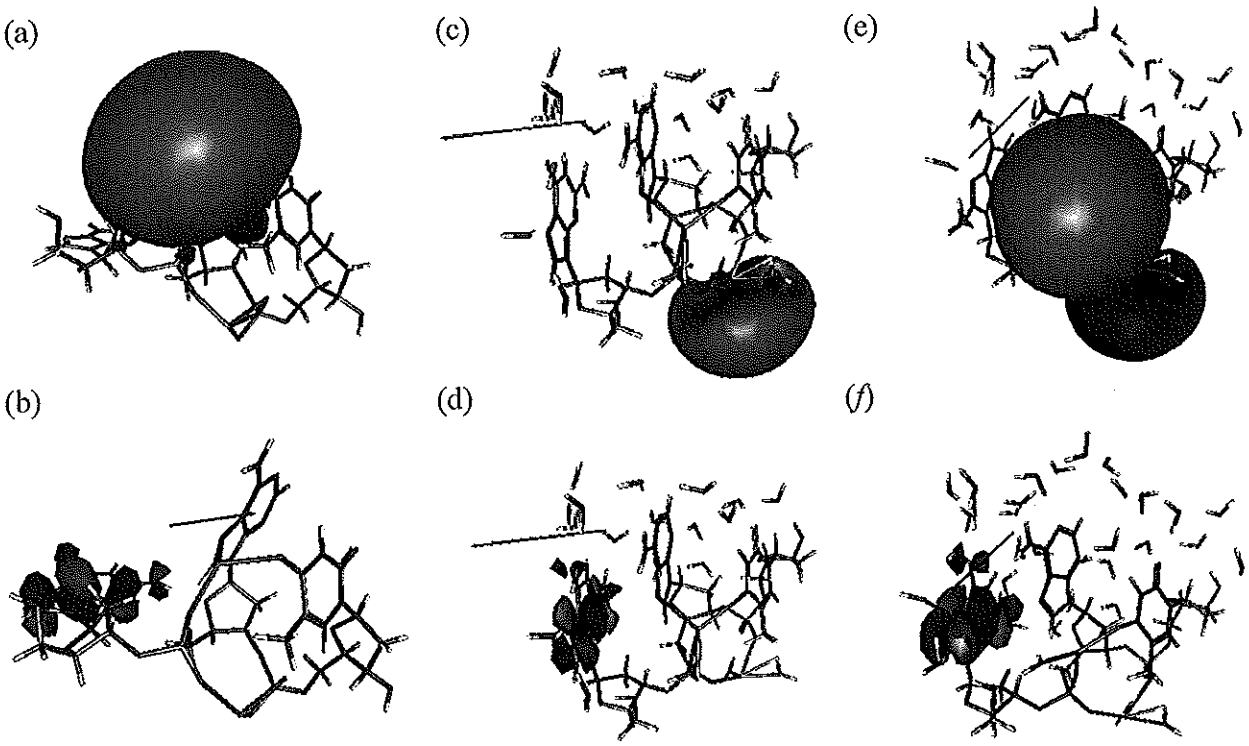


Fig. 16 (a), (c), and (e) show the LUMO states of the codon 5'-GAT-3' without water molecules, with 20 water molecules, and with 33 water molecules, respectively. (b), (d), and (f) show the HOMO states of the codon 5'-GAT-3' without water molecules, with 20 water molecules, and with 33 water molecules, respectively.

The main calculation results are listed below.

1) Influence of the water molecules on the crystal structures of the codon GAT When water molecules are included in the simulation, the crystal structure of the codon changes from the unclusterized structure to that aggregated on the backbone where the bases tends to parallel to each other. The number of water molecules greatly influences the direction of the dipole moment of the codon, which means that the IR spectrum of the codon will be changed.
2) The water molecules are mainly localized next to the hydrogen molecules of the molecules. There is no water molecules in between the nucleotide bases and around the backbone of the codon. Furthermore, there is no water molecules at the ends of the codons.

3) The water molecules do not change the localization of the HOMO state of the codon. However, they do change the LUMO state of the molecule. Here, as we can seen in the figure, the water molecules relocate the state to the backbone and extend the state to the A base. This action of the system will reduce the energy gap between the LUMO state and the HOMO state.

The analysis of the numerical calculation results of the influence of water molecules on the energy structures of ssDNAs and the vibrational characteristics of ssDNAs is in progress.

4. Chaotic and Ultrahigh Frequency Rabi Oscillation in a Coupled-Double-Quantum-Dot Semiconductor System

The mechanism of the charge transfer inside a DNA molecule under the influence of THz radiation is crucially important to the study of the relationship between the IR/THz modulated current and the sequence of the nucleotides. As well accepted, the G DNA plays a role of quantum well and T DNA acts as a potential barrier in the process of charge transfer inside a DNA molecule. To this connection, in terms of density matrix theory, we have studied the Rabi oscillation of the charge transfer in a coupled-double-quantum-dot system under a high THz field. The research results are not only suitable to semiconductor system but also to DNA molecular system. For example, a DNA molecule with sequence of ATGGTGGTA is a counterpart of the semiconductor coupled quantum dot system. The main results are listed below.

1) The probability oscillation exhibits chaotic features as well as the characteristics of collapse and revivals of the oscillations. Here, the nonlinear characteristic is due to the multi-photon excitations of the electron.
2) The coherence time of the electron may be greatly extended through the collapse and revival mechanism of the oscillations. The chaos-enhanced coherence may have important application in quantum information processing.
3) The creation of the ultrahigh frequency oscillation of the electronic probability in each probability step. This characteristics of the oscillation may have important application in ultrafast quantum information processing.

This paper has been submitted to Physical Review Letters. The reviewers of the journal presented some comments. The revision of the paper is in progress.

III. On-going Research and Future Work

1. Investigation of the conduction mechanism of charge in DNA molecules.

In our current research, we found that the number of the conductive channels of the electrons in a DNA molecule and the magnetite of the transmission coefficients is greatly related to the length of or the number of bases in the molecule. More interestingly, the conduction characteristics of the charge verse the length of the molecule is not a monotonic function. The figure below shows the relationship between the transmission coefficients of the charge and the number of bases clearly. This result is very important to information transfer in DNA/RNA and in bio-electronics. It should be noted that the result is obtained in terms of the tight binding theory which ignores the detail of the interaction between the nucleotides and that between the bases and the environment. Currently, we are studying this issue in terms of density functional theory and Green's function theory for transport of charges inside a DNA molecule. This study will reveal the answer to this issue and predict relationship between the conductivity and the length of the molecule and the environment.

2. IR/THz absorptions of larger DNA/RNA molecules

As we have discussed above, the IR/THz absorption of a DNA molecule is of importance to label-free-optical-related DNA sequencing techniques. In previous research, we have studied the characteristics of IR/THz absorption of DNA codons and found the general rules for the identification of the codons in gas phase. Recently, we finished the geometrical optimization of the ssDNA with three codons for the study of the influence of the interaction among codons on the absorption characteristics of the molecule. Specified research topics in this research direction include:

1) Influence of water molecules on the IR/THz absorption spectra of genetic codons.
2) The IR/THz absorption of DNA/RNA molecules with 4 ~12 bases.
3) Molecular dynamics based investigation of the vibrational spectra and IR/THz absorption spectra of DNA/RNA molecules
4) Molecular dynamics based investigation of the vibrational spectra and IR/THz absorption spectra of DNA/RNA molecules bonded on the surface of electrodes

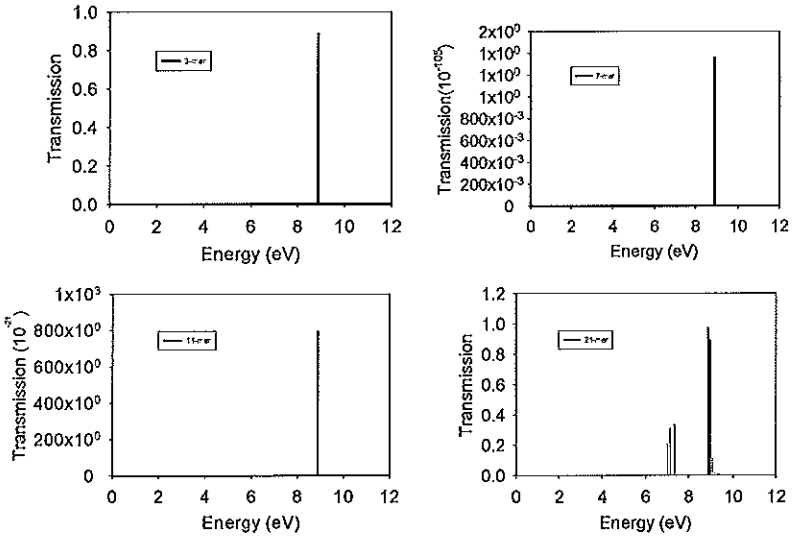


Figure 17 Relationship between the transmission coefficients of charges through DNA molecules and the number of nucleotides inside the molecules.

3. Relationship between IR/THz modulated current and the sequence of nucleotides in DNA/RND molecules

In task 1, we focus our research attention on the understanding to the mechanism of conduction of charge in short length DNA molecules (with four to fifteen bases). In this research task, we will focus on the relationship between the IR/THz modulated current and the sequence of nucleotides in the molecules. Here, special attention will be focused on the following issues.

1) Sequence dependent and IR/THz modulated current through a single DNA/RNA molecule. This research aims at the relationship between the current and the sequence of the nucleotides in the molecules.
2) How to increase the absorption of the radiation for the detection of the sequence information through the measurement of the current through the molecule.
3) Sequence dependent and IR/THz modulated current through a single DNA/RNA molecule bonded on the surface of electrodes.

IV. Publications and Conference Presentations

1. Submitted Papers

1) Chaotic and Ultrahigh Frequency Rabi Oscillation in a Coupled-Double-Quantum-Dot Semiconductor System, Submitted to Physical Review Letters; Revision of the letter is in progress.

Abstract: The chaotic and ultrafast oscillatory characteristics of the probability of the electron in a coupled-double-quantum-dot-based two-level system are studied in terms of density matrix theory. The characteristics are due to multi-photon excitations of the electron and the related coherent nonlinearity of the system. The chaos-enhanced coherence and the high frequency characteristics of the oscillations may have important application to ultrafast quantum information processing.

2) Sequence Dependence of Vibrational Characteristics of Genetic Codons (submitted to Applied Spectroscopy)

Abstract: A theoretical study of the vibrational characteristics of all DNA codons in gas phase is presented. The simulation results show that the spectra of the codons in the high frequency end of the middle infrared regime can be classified into five semi-distinct spectral sub-regimes, namely: two end regions and one region that separates the two sequence identifier regions with features that arise primarily from coupling between the GG bases and other bases (T, C, and A). This study reveals a number of phenomenological trends that have scientific relevance to spectroscopic characterization of DNA molecules. For example, spectra in the high-frequency end-regions arise primarily from vibrational modes associated with the O-H bonds at the 5 ' end or 3 ' end of the codons and the spectra in the low-frequency end-regions are formed primarily by the vibrational modes of the backbone. Conversely, in the lower-frequency side, the sequence-identifier region is composed of the C-H bond stretch vibrations in the planes of the

corresponding DNA bases, and in the higher-frequency side, sequence-identifier region is composed of the N-H bond stretch vibrations in the planes of the corresponding DNA bases. In addition, the sequence-identifier dividing region almost exclusively contains vibrational modes due to coupling between the G nucleotides and other bases. As will be shown, all the vibrational modes in sequence identifier regions are localized at the corresponding DNA bases and exhibit a definable dependence on the sequence form of the codons under study. In addition, the simulation results also show that all the vibrational modes in sequence identifier regions (high frequency end) shows the following features: a) The spectral lines at high frequency end of the spectrum show "two group feature"; b) There is a frequency forbidden region in the region with some exceptions; c) The number of spectral lines increases while the nucleotide at the third place of a codon changes from T to G; d) The number of spectral lines in the first frequency identifier region satisfies the relation: if the number of T bases in a ssDNA is N_T, the number of C bases or A bases in a ssDNA is N_C or N_A, and the number of G bases in a ssDNA is N_G, the total number of spectral lines in first frequency identifier region is

$$N = N_T + 2(N_C+N_A) + N_G$$

2. Papers Finished and To be Submitted

1) Sequence dependence of the energy structures of DNA codons (finished, to be submitted to Physical Review E)

Abstract: We have theoretically studied the sequence dependence of the energy level structures of the genetic codons in gas phase. The calculated results show that the sequence of nucleotides in a DNA codon can greatly influence the HOMO and LUMO energies of the codons. If the third nucleotide in a codon is changed from Pyrimidine bases to purine bases, the change of the energy difference between the LUMO and the HOMO energies of the codons is big ranging from 0.01 eV to 0.03 eV which depends on the structure of the first two nucleotides. If the codons have the similar sequence, like GGA AGG, the interaction of the codons produce a large number of energy levels next to the zero energy of the system. The change in sequence of the second codon, for example GGA AGT, could greatly influence the energy structure of the system.

2) Influence of Codon Interactions on the Vibrational Characteristics of ssDNAs (finished, to be submitted to Physical Review E)

Abstract: We have theoretically studied the influence of DNA codons on the sequence dependence of the absorption spectra of ssDNA molecules in gas phase in high frequency end of middle infrared regime 3000 cm^{-1} ~ 4000 cm^{-1}. The calculated results show that the influence of the codon interaction on the absorption spectra of the ssDNA molecules can be expressed in the following aspects: all vibrational modes in the sequence identifier regions are localized at the corresponding DNA bases, which is crucially important to optical-based label free DNA sequencing techniques; A new group of spectral lines is created around 2940 cm^{-1}; The intensities of the IR absorptions of all vibrational modes is reduced by the interaction; The number of vibrational modes increases due to the interactions; In short wavelength Mid-IR

region, the interaction between DNA codons eliminates the C-H bond vibrational modes which are localized in the corresponding nucleotides.

3) Influence of Codon Interactions on the Energy Structures of ssDNAs (finished, to be submitted to Physical Review E)

Abstract: We have theoretically studied the influence of the interaction DNA codons on the energy structures of ssDNA molecules in gas phase. A multilevel interaction model is presented for the explanation of the simulation results. The calculated results show that the influence of the codon interaction on the energy of the ssDNA molecules is strongly sequence related. Here, the Interaction between GGA and AGG may greatly reduce the energy difference between the LUMO state and the HOMO state of the ssDNA. In contrast, the interaction between GGA and AGA doesn't change the energy difference between the LUMO and the HOMO states of the codon GGA too much. Generally, for the ssDNA 5'-GGA XYZ-3', the energy difference between the LUMO and the HOMO states is the smallest when X = T.

4) Influence of Codon Interactions on the Energy Structures of ssDNAs (finished, to be submitted to Physical Review Letters or Physical Review E)

Abstract: A dipole-dipole interaction model and a graphical analysis technique is presented to qualitatively explain the crystal structure of single stranded DNA (ssDNAs). Here, DNA codons, which carrying the hereditary information for protein synthetics, are the fundamental units or blocks of the crystal structures of ssDNAs. Hence, the electric interaction between the dipoles of adjacent codons determines the overall form of the ssDNA while the base paring interaction modifies the geometrical parameters. Furthermore, the H-bone interaction between bases on the backbone and the Na ion mediated base pairing interaction determine the aggregation of the bases on the backbones of ssDNAs

V. References

[1] FS Collins, ED Green, AE Guttmacher, and MS Guyer , "A vision for the future of genomics research", Nature 422, 835–847(2003).

[2] Lloyd M. Smith, Jane Z. Sanders, Robert J. Kaiser, Peter Hughes, Chris Dodd, Charles R. Connell, Cheryl Heiner, Stephen B. H. Kent , and Leroy E. Hood, "Fluorescence detection in automated DNA sequence analysis", Nature 321, 674–679(1986).

[3] JM Prober, GL Trainor, RJ Dam, FW Hobbs, CW Robertson, RJ Zagursky, AJ Cocuzza, MA Jensen, and K Baumeister, "A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides", Science 238, 336–341(1987).

[4] J. Ju, C. Ruan,C. W. Fuller, A. N. Glazer, R. A. Mathies, "Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis", Proc Natl Acad Sci USA, 92, 4347–4351(1995).

[5] Cheuk-Wai Kan, Erin A. S. Doherty, Annelise E. Barron, "A novel thermogelling matrix for microchannel DNA sequencing based on poly-N-alkoxyalkylacrylamide copolymers", Electrophoresis 24:4161–4169.

[6] Snezana Drmanac, David Kita, Ivan Labat, Brian Hauser, Carl Schmidt, John D. Burczak & Radoje Drmanac, "Accurate sequencing by hybridization for DNA diagnostics and individual genomics", Nat Biotechnol 16, 54–58 (1998).
[7] Dong-Jing Fu, Kai Tang, Andreas Braun, Dirk Reuter, Brent L. Iverson, Brigitte Darnhofer-Demar, Daniel P. Little, Maryanne J. O'Donnell, "Sequencing exons 5 to 8 of the p53 gene by MALDI-TOF mass spectrometry", Nat Biotechnol 16, 381–384 (1998).
[8] M. T. Roskey, P. Juhasz, I. P. Smirnov, E. J. Takach, S. A. Martin, L. A. Haff , "DNA sequencing by delayed extraction-matrix-assisted laser desorption/ionization time of flight mass spectrometry", Proc Natl Acad Sci USA, 93, 4724–4729 (1996).
[9] Edwards JR, Itagaki Y, Ju J (2001) Nucleic Acids Res 29:E104–E104.
[10] John J. Kasianowicz, Eric Brandin, Daniel Branton, and David W. Deamer, "Characterization of individual polynucleotide molecules using a membrane channel", Proc Natl Acad Sci USA 93, 13770–13773 (1996).
[11] Jay Shendure, Gregory J. Porreca, Nikos B. Reppas, Xiaoxia Lin, John P. McCutcheon, Abraham M. Rosenbaum, Michael D. Wang, Kun Zhang, Robi D. Mitra, and George M. Church, "Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome", Science 309, 1728–1732(2005).
[12] M. Ronaghi, M. Uhlen, P. Nyren, "A Sequencing Method Based on Real-Time Pyrophosphate", Science 281, 363–365(1998).
[13]Ido Braslavsky, Benedict Hebert, Emil Kartalov, and Stephen R. Quake, "Sequence information can be obtained from single DNA molecules", Proc Natl Acad Sci USA, 100, 3960–3964 (2003).
[14] R. D. Mitra, J. Shendure, J. Olejnik, O. Edyta Krzymanska , G. M. Church, "Fluorescent in situ sequencing on polymerase colonies", Anal Biochem 320:55–65(2003); John Eid, et al, "Real-Time DNA Sequencing from Single Polymerase Molecules", Science, 323, 133-138(2008).
[15] DL Woolard, R Brown, M Pepper, M Kemp, "Terahertz frequency sensing and imaging: a time of reckoning future applications?", Proceedings of the IEEE, 93(10), 1722-1743(2005).
[16] Tatiana Globus, Dwight Woolard, Thomas W Crowe, Tatyana Khromova, Boris Gelmont, Jeffrey Hesler, "Terahertz Fourier transform characterization of biological materials in a liquid phase", J. Phys. D: Appl. Phys. 39, 3405(2006).
[17] James D. Waston, "Molecular Biology of the Gene", third edition, W. A. Benjamin, Inc. (1976, Philippines).
[18] Yamuna Krishnan and Friedrich C. Simmel,"Nucleic Acid Based Molecular Devices", Angewandte Chemie International Edition, Volume 50(14),3124–3156(2011)
[19]. Schneider, B., Cohen, D. & Berman, H. M. (1992) Biopolymers 32, 725–750; Halle, B. & Denisov, V. P. (1998) Biopolymers 48, 210–233; 5. Umehara, T., Kuwabara, S., Mashimo, S. & Yagihara, S. (1990) Biopolymers 30, 649–656; Duan, Y., Wilkosz, P., Crowley, M. & Rosenberg, J. M. (1997) J. Mol. Biol. 272, 553–572.
[20]. Hess, S., Davis, W. B., Voityuk, A. A., Rosch, N., Michel-Beyerle, M. E., Ernsting, N. P., Kovalenko, S. A. & Lustres, J. L. P. (2002) Chemphyschem. 3, 452–455; Pal, S. K., Peon, J. & Zewail, A. H. (2002) Chem. Phys. Lett. 363, 57–63.
[21]. Latt, S. A. & Stetten, G. (1976) J. Histochem. Cytochem. 24, 24–33; Gorner, H. (2001) Photochem. Photobiol. 73, 339–348.